EDA Project

Tabea Nöthe-Menchen

The Task

- King County Dataset of houses and sales
- total of 21,597 houses
- 2.5 days for analysis and preparing summary

The Client

Larry Sanders

The client has the following requirements for his new home:

- it should be located at the **water**

- Larry has a **limited budget** available

- the house should be in a **nice** condition

- the house should be located rather **isolated**

- **no children** should live in the immediate **neighborhood**, because Larry is afraid of germs

- he himself has **some children**

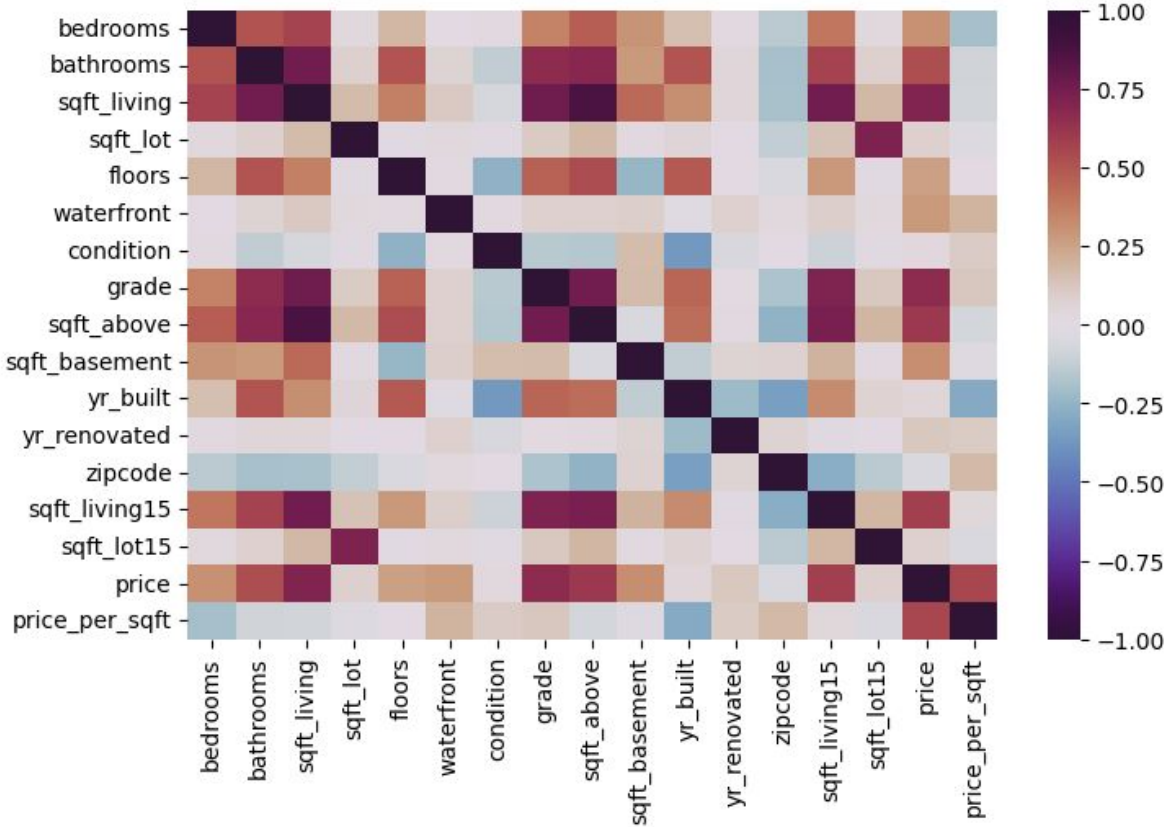
Questions | Hypotheses

- Does the requirement ‘waterfront’ affect the price of the potential houses?
  → on average houses at the waterfront are more expensive

- Is the price of a house associated with its condition?
  → the better the condition of the house the more expensive it is

- Can the size of the living area be used as a predictive number for the number of bedrooms?
  → The higher the number of bedrooms the greater the square footage of the living area

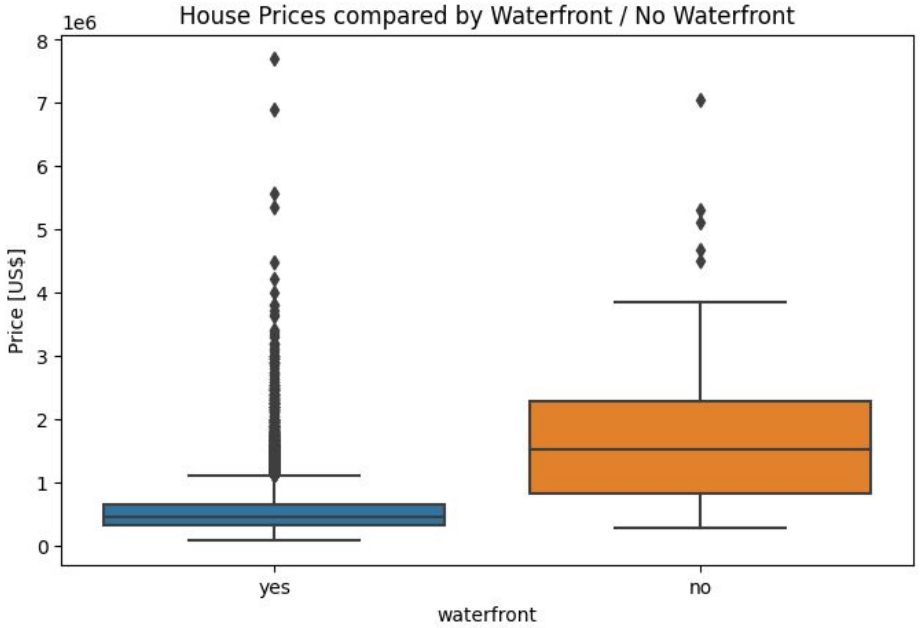
Correlations in all Data



‘waterfront’ is positively correlated with the price

146 properties are located at the waterfront.

	mean price	mean price/sqft
waterfront	1,717,214.73 $	529.23 $/sqft
no waterfront	532,771.21 $	256.91 $/sqft



Pear’s Correlation Coefficient: 0.282

How many bedrooms?

- Larry has ‘some’ children

Assumptions:

- Larry has **at least 2** children
- most people would like to have **a bedroom for each of their kids**

→ the house should have **at least 3 bedrooms**

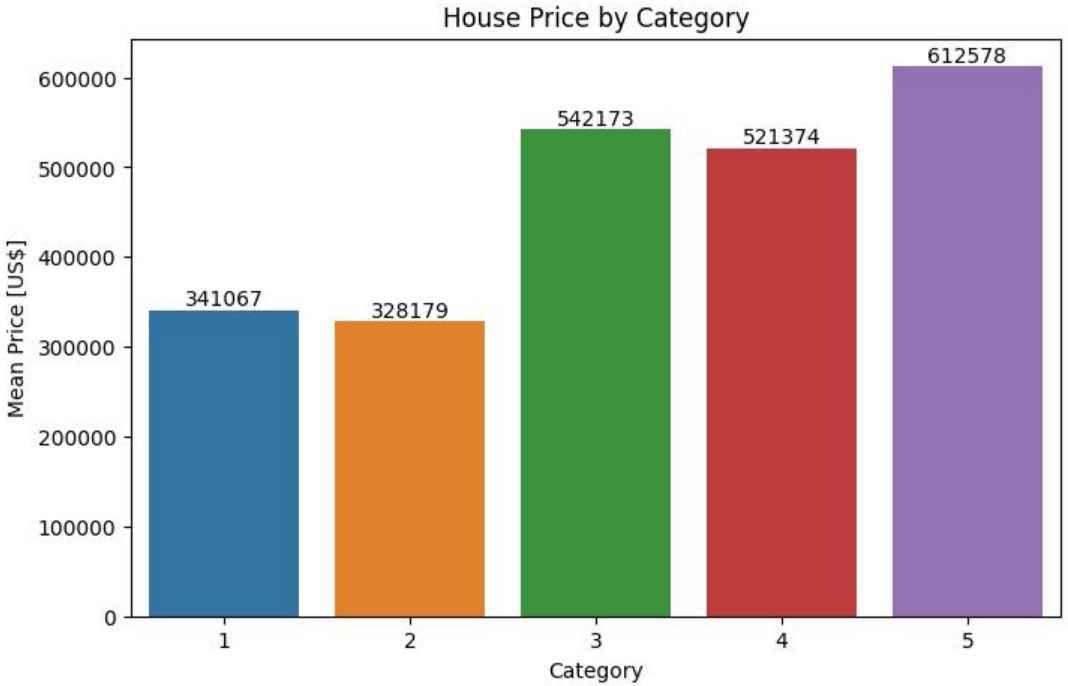
114 objects have at least 3 bedrooms.
(average price = 1,946,065.35 $)

Is the price of a house associated with its condition?

Pear's Correlation Coefficient: 0.033

→ The price **is not strongly correlated** with the 'condition' of the property.

46 properties meet the given criteria.
('condition' > 3)

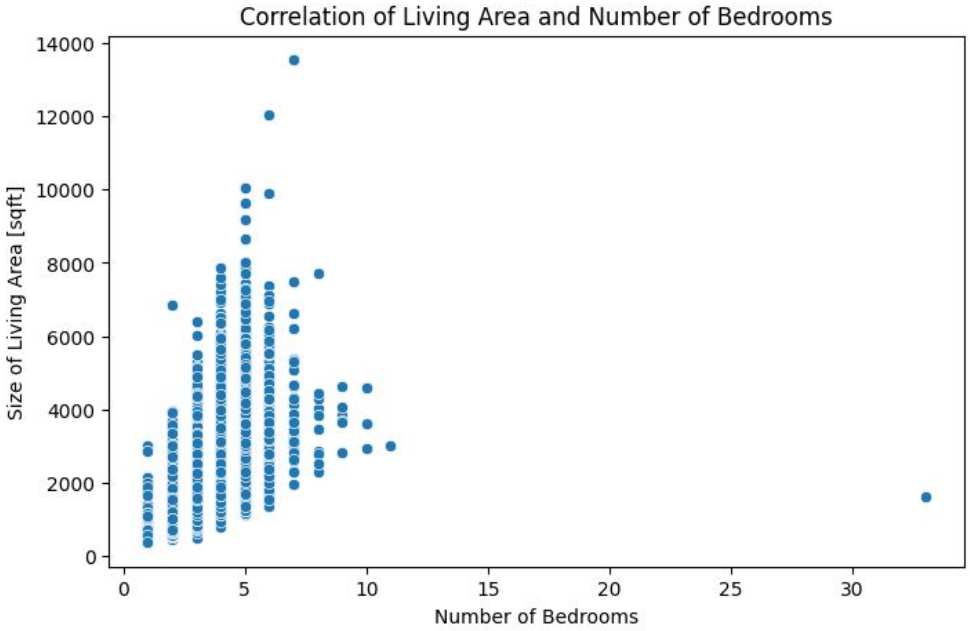


Correlation of number of bedrooms and living area

- the house should be located rather **isolated**
- **no children** should live in the immediate **neighborhood**, because Larry is afraid of germs

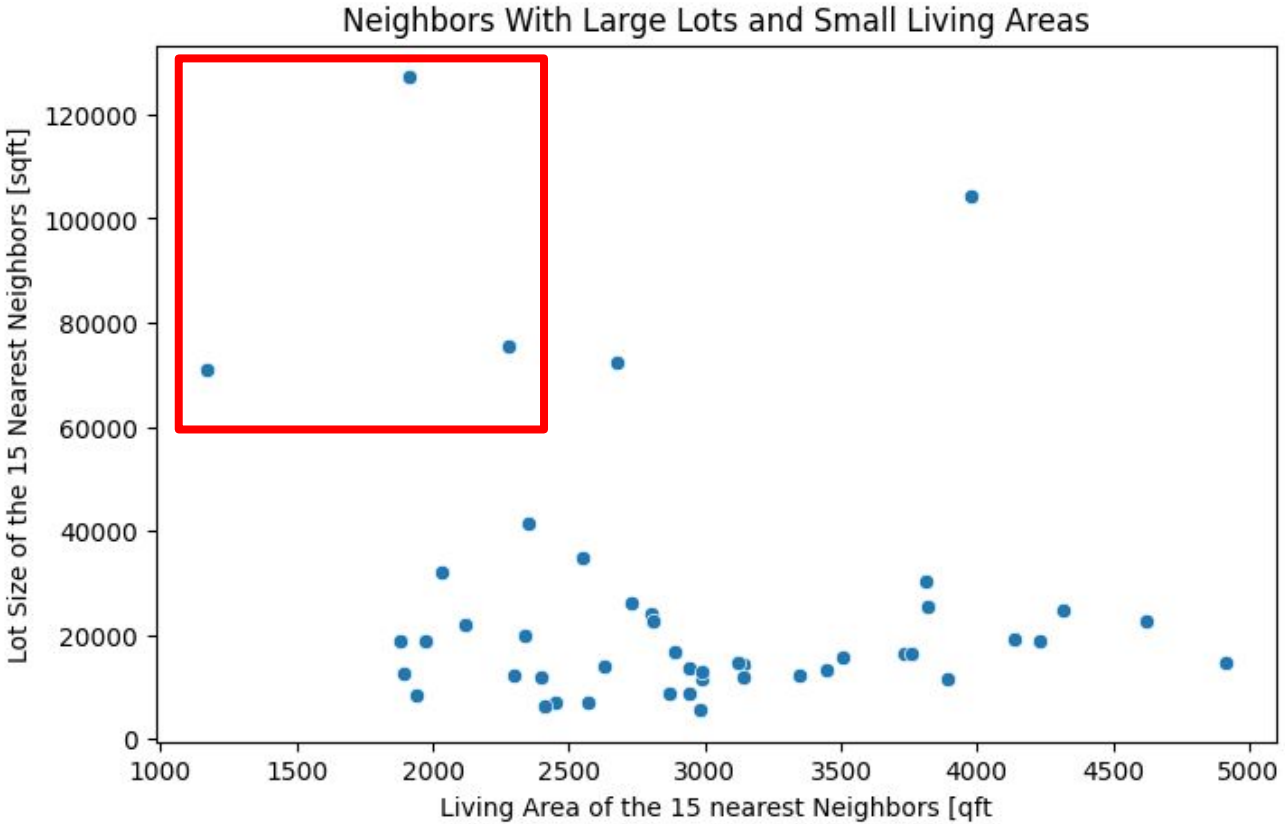
- correlation of number of bedrooms and size of living area
  → mean living area of neighbors could be used as an indicator of the number of bedrooms in the neighborhood and thus as an indicator of presence or absence of children

- an house is rather isolated if the neighbors have large lots



Pear's Correlation Coefficient: 0.572

Which properties meet criteria the best?



suggestion 1

- waterfront
- at least 3 bedrooms
- at least condition rank 4
- `sqft_living15` small value
- `sqft_lot15` large value
- lowest price

# bedrooms	3
# bathrooms	1.75
living area [sqft]	1940
lot [sqft]	167125
waterfront	yes
condition	4
grade	7
year built	1955
living space of neighbors [sqft]	1910
lot area of neighbors [sqft]	127,195
price [$]	635,000
price per sqft [$/sqft]	327.32

suggestion 2

- **NO** waterfront
- at least three bedrooms
- condition at least rank 4
- sqft_living15` small value
- `sqft_lot15` large value

# bedrooms	3
# bathrooms	2.75
living area [sqft]	2290
lot [sqft]	34548
waterfront	no
condition	4
grade	7
year built	1984
living space of neighbors [sqft]	399
lot area of neighbors [sqft]	275,299
price [$]	536,000
price per sqft [$/sqft]	234.06

suggestion 3

- waterfront
- at least three bedrooms
- condition at least rank 3
- lowest price

# bedrooms	4
# bathrooms	1
living area [sqft]	1200
lot [sqft]	11834
waterfront	yes
condition	3
grade	6
year built	1972
living space of neighbors [sqft]	1670
lot area of neighbors [sqft]	47,462
price [$]	340,000
price per sqft [$/sqft]	283,33

	#1	# 2	# 3
# bedrooms	3	3	4
# bathrooms	1.75	2.75	1
living area [sqft]	1940	2290	1200
lot [sqft]	167125	34548	11834
waterfront	yes	no	yes
condition	4	4	3
grade	7	7	6
year built	1955	1984	1972
living space of neighbors [sqft]	1910	399	1670
lot area of neighbors [sqft]	127,195	275,299	47,462
price [$]	635,000	536,000	340,000
price per sqft [$/sqft]	327.32	234.06	283,33

Location and Price of Suggested Houses